Caltech Library

Teaching 21st Century Knowledge Management Skills: A Carpentry Approach to Open Data, Software, and Publication

January 19, 2017
1:00 - 2:15 PM | 2:30 - 3:45 PM
Sherman Fairchild Library, Multimedia Conference Room (Room 328)

Gail Peretsman-Clement (gperetsm@caltech.edu)
Donna Wrublewski (dtwrub@caltech.edu)

Overview

> "Young investigators are still learning an onslaught of professional skills…
> They have less time than ever to do research or ponder big ideas."
>
> "Early-career researchers need fewer burdens and more support,"
> *Nature* Editorial, **538**, p. 427, October 27, 2016. doi:10.1038/538427a

Today's researchers are challenged to master an ever-expanding and interlinked set of computational, data management, authorship and publishing skills in the rapidly evolving scholarly web. An emerging approach, called Carpentry, teaches researchers at all career stages best practices and efficient tools for producing, documenting, handling, sharing, publishing, and providing fair attribution for research outputs of all types: data, software, and papers. The ultimate aim of software, data and author carpentry is to support more open, transparent, and reusable research to advance and strengthen science in the digital era.

This workshop will introduce you to principles of Carpentry, teach you a Data Carpentry module on data cleaning using OpenRefine, and facilitate a debriefing discussion about the hands-on, technology-enabled, open-source and collaborative teaching methods you will have experienced in the demo.

software carpentry — https://software-carpentry.org/

DATA CARPENTRY — MAKING DATA SCIENCE MORE EFFICIENT — http://www.datacarpentry.org/

AuthorCarpentry — http://libguides.caltech.edu/authorcarpentry

OpenRefine — http://openrefine.org/

Exercise 1: Watch instructor demo of OpenRefine's Faceting and Clustering features. Using individual laptop, practice these skills on data values in the column 'Institution'

Notes:

Resources needed:

- Dataset (downloaded to desktop): nobel_prize_carpentry_lesson.csv
- OpenRefine session running on Chrome or Firefox, with dataset open
- Pink and blue stickies to signal 'All okay' or 'Help needed'
- GoogleDoc named 'TeachWeek Shared Notes' for typing questions or comments

Exercise 2: In pairs, apply Faceting and Clustering to the data in the column 'BornCity' (Group A) or 'BornCountry' (Group B.)

Group A. Which are the top three cities that have produced Nobel Laureates? Did Faceting and Clustering help lead to this answer? Why or Why not? (*Please record your answers in the Google Doc. Instructor will request one pair to demo their results on the black monitor using AirMedia app*).

Group B. Which are the top 5 countries that produced Nobel Laureates? Did Faceting and Clustering help lead to this answer? Why or Why not? (*Please record your answers in the Google Doc. Instructor will request one pair to demo their results on the black monitor using AirMedia app*)

Notes:

Resources needed:

- Dataset (downloaded to desktop): nobel_prize_carpentry_lesson.csv
- OpenRefine session running on Chrome or Firefox, with dataset open
- Pink and blue stickies to signal 'All okay' or 'Help needed'
- GoogleDoc named 'TeachWeek Shared Notes' for typing answers and questions
- AirMedia app to project OpenRefine results to black wall monitor

Exercise 3a: Work in pairs to calculate the number of awardees that were born in present-day Germany. Does Germany overtake the UK in terms of awardees?

Exercise 3b: Without consulting OpenRefine, design a search strategy that would automate this calculation.

(***Please record your answers in your space on the whiteboard. Group will review all answers together***)

Resources needed:

- Dataset (downloaded to desktop): nobel_prize_carpentry_lesson.csv
- OpenRefine session running on Chrome or Firefox, with dataset open
- Pink and blue stickies to signal ‘All okay’ or ‘Help needed’
- Whiteboard markers and some whiteboard real estate

Additional readings and resources

Sherman Fairchild Library Multimedia Conference Room (MCR) Specifications

- Multimedia System: Crestron 3-Series® DigitalMedia™ Presentation System 300 w/Audio Conferencing Interface, model DMPS3-300-C-AEC
    - https://www.crestron.com/products/model/DMPS3-300-C-AEC
    - AirMedia allows wireless display of up to 4 devices simultaneously (AM-101 System): https://www.crestron.com/products/model/am-101
- Projector: EPSON BrightLink Pro 1430Wi Collaborative Whiteboarding Solution with Touch Interactivity
    - https://epson.com/For-Work/Projectors/Interactive/BrightLink-Pro-1430Wi-Collaborative-Whiteboarding-Solution-with-Touch/p/V11H665520
- Monitor: Sharp Aquos LE-90" CLASS LED SMART 3D TV, model LC-90LE745U
    - http://www.sharpusa.com/ForHome/HomeEntertainment/LCDTV/Archives/LC90LE745U.aspx
- Webcam: Panasonic FreeTalk Conference HD Camera, Model TALK-7181
    - http://shop.panasonic.com/support-only/TALK-7181.html
- Lectern: LEX30 High Tech Multimedia Lectern
    - http://www.video-furn.com/podiums-lecterns-LEX30.html?v=%3C?php%20echo%20$_GET%5B%27v%27%5D;%20?%3E
- Whiteboard paint: FUZE Dry Erase Paint
    - https://www.fuzewall.com/order
- Power Towers: iTeach Mobile Power Tower Power Strip, with USB power ports
    - https://moorecoinc.com/feature-sheets/iteach-mobile-power-tower.pdf
    - http://www.todaysclassroom.com/balt-27735-iteach-mobile-power-tower/
- Additional supplies: Dry erase markers and erasers, dry erase board cleaner, towels

Carpentry Background Reading

- Wilson G. Software Carpentry: lessons learned [version 2; referees: 3 approved]. F1000Research 2016, 3:62 (doi: 10.12688/f1000research.3-62.v2)
    - https://f1000research.com/articles/3-62/v2
- Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, Pawlik A. Data carpentry: workshops to increase data literacy for researchers. International Journal of Digital Curation. 2015 Mar 18;10(1):135-43.
    - http://dx.doi.org/10.2218/ijdc.v10i1.351

OpenRefine

- OpenRefine Documentation: http://openrefine.org/documentation.html
- *Using OpenRefine* (ebook): https://caltech.tind.io/record/903269
- Clustering In Depth - OpenRefine Wiki: https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth
- GREL String Functions: https://github.com/OpenRefine/OpenRefine/wiki/GREL-String-Functions

Dataset Source

- Modified from Nobelprize.org Developer Zone
    - https://www.nobelprize.org/nobel_organizations/nobelmedia/nobelprize_org/developer/